Studies in History and Philosophy of Modern Physics 38 (2007) 732–756

Studies in History and Philosophy of Modern Physics

www.elsevier.com/locate/shpsb

The problem of time's arrow historico-critically reexamined

Roberto Torretti

Department of Philosophy, University of Puerto Rico, P.O. Box 21572, San Juan, Puerto Rico 00931-1572, USA

Received 13 August 2006; received in revised form 12 November 2006; accepted 20 November 2006

Abstract

Responding to Hasok Chang's vision of the history and philosophy of science (HPS) as the continuation of science by other means, I illustrate the methods of HPS and their utility through a historico-critical examination of the problem of "time's arrow", that is to say, the problem posed by the claim by Boltzmann and others that the temporal asymmetry of many physical processes and indeed the very possibility of identifying each of the two directions we distinguish in time must have a ground in the laws of nature. I claim that this problem has proved intractable chiefly because the standard mathematical representation of time employed in the formulation of the laws of nature "forgets" one of the connotations of the word 'time' as it is used in ordinary language and in experimental physics.

Keywords: Time; Direction of time; Entropy; Boltzmann entropy; Big Bang and entropy

1. The continuation of science by other means

Hasok Chang's (2004) book *Inventing Temperature* offers in its last chapter a view of the history and philosophy of science (HPS) as "the continuation of science by other means". The need for supplementing normal scientific research in this fashion was impressed on Chang by his personal experience as a philosophical historian of science, which he describes as "a curious combination of delight and frustration, of enthusiasm and skepticism, about science". His delight in the beauty of conceptual systems and the

E-mail address: roberto.torretti@yahoo.com.

1355-2198/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsb.2006.11.005

masterfulness of experimental setups is mixed with "frustration and anger at the neglect and suppression of alternative conceptual schemes, at the interminable calculations in which the meanings of basic terms are never made clear, and at the necessity of accepting and trusting laboratory instruments whose mechanisms" he does not understand (p. 236). Chang claims that HPS can actually *generate* scientific knowledge in at least two ways. On the one hand, through the recovery of forgotten scientific knowledge, HPS can reopen neglected paths of inquiry. On the other hand, by applying the philosopher's scalpel to the thick and often opaque tissue of scientific discourse, HPS can positively contribute to clarifying or eliminating the confused, ambiguous or downright inept notions that bedevil innovative scientific thinking and appear to blunt its cutting edge.

Encouraged by Chang's proposal, I offer here a historico-critical examination of the problem that A. S. Eddington labeled with the catch phrase "the arrow of time".[1] This problem is still the subject of impassioned arguments and few would pronounce it closed. Its persistence is due, in part, to the strong emotions and *weltanschauliche* commitments associated with the word 'time'; but it also owes much to the common inclination to identify the broad ordinary meaning of this word with the special meaning it takes in some scientific contexts, or—worse still—to assume that the streamlined "technical" meaning is somehow better or truer than the ordinary one, some of whose connotations it lacks.

2. The problem of time's arrow

The phrase "time's arrow" was coined by Eddington (1929, p. 69) presumably on the analogy of the arrows placed on street corners to indicate the direction of traffic. The idea that time flows—where? in another time?—is senseless, but it is old and popular. In an otherwise beautiful line, Vergil (*Georg.* 3.284) wrote that "time flies away", without bothering to specify the medium in which this feat occurs. In the professional literature about time's arrow, the expression does not usually refer to an attribute *of* time, but rather to the patterns of succession of natural events *in* time.[2] The strange impression we get from watching a videotape—e.g. of a football game—while it is being rewound would indicate that common physical processes follow patterns of occurrence that normally cannot be reversed. From this standpoint, Eddington's phrase might suggest that time itself sets a direction or order to events. However, the literature concerning time's arrow, besides gathering and describing patterns of succession that appear to be irreversible, generally seeks to explain their irreversibility as a consequence of the universal laws of nature. Such attempts must overcome one major difficulty. We normally assume that the fundamental equations of classical, relativistic and quantum mechanics and electrodynamics express the universal laws of nature to an approximation that is sufficiently good in their respective

[1]This paper is based on Section 4 of a much longer paper entitled "Can science advance effectively through philosophical criticism and reflection?", which I deposited on 13 August 2006 at http://philsci-archive.pitt.edu. The text has been revised in the light of the Editors' comments, two referee reports and private communications by Olimpia Lombardi and Hasok Chang. I am glad to acknowledge here their valuable advice and to thank them for it, while assuming full responsibility for the errors that remain.

[2]Gell-Mann & Hartle (1994), p. 311, Halliwell, Pérez-Mercader, and Zurek (1994) list six different such patterns or "arrows of time". Surprisingly, they fail to mention the one that most closely concerns us humans, and which may well be mainly responsible for our infallible sense of temporal orientation: we start living at birth and thereupon grow bigger and older until we finally die; not a single case is known of a person who rose from the grave and thereupon grew younger and ended by climbing up into his or her mother's womb.

fields of application. Those equations are invariant under the *time reversal* transformation $t \mapsto -t$, which multiplies every value of the time variable by -1. This invariance implies that for every temporal series of phenomena represented by a solution of the equations there is a matching series represented by another equally true solution, in which the corresponding phenomena succeed one another in reverse order. In many cases, however, only one of each such pair of solutions is exemplified in the natural world. To explain this selectivity of nature by deriving it from time reversal invariant laws is an ambitious undertaking which, to say the least, is not very likely to succeed.

3. Meanings of 'time'

When faced with the problem, a philosophically minded person will ask, in the first place, what the word 'time' means and how it is being used. Since 'time' is a noun it is plausible to ask for its referent. Reading some philosophers one even gets the impression that the word designates a unique entity and therefore ought to be regarded as a proper name (although in English we seldom capitalize it). Its purported *denotatum* is, however, hard to pin-point. In everyday conversation, 'time' is most frequently employed as a common noun, to denote particular instants or particular durations. Kant held that, in stark contrast with ordinary common nouns, the relation between the several objects called 'times', in the plural, and that which we call 'time', in the singular, is not that of individual *instances* to the *class* to which they belong, but rather that of *parts* to a *whole.* Yet, if the whole of Time remains elusive, the most one can grant to Kant is that, among those multiple items to which the common noun 'time' refers, some are related to each other as parts to wholes, while others—viz., the instants—are related to the former as their boundaries. Something like this is probably what most of us would come up with when prompted to elucidate 'time'. Still, I do not think that 'time' must denote an individual object or a class of such objects merely because it performs like a noun. On the other hand, I see no difficulty in taking 'time' as a portmanteau term that connotes a wide variety of aspects of our life in the world, but does not denote anything in particular. If such is the case, those who ask "What is time?" should not expect a simple, non-contextual, reply.

For our present purposes, it will be sufficient to consider four pervasive features of our human experience that give good reason for describing it as experience *of* time and *in* time. I designate them as *waiting times*, *time points* (or *instants*), *time order*, and the threefold display of *past, current and future times* at each actual instant. I have deliberately included the word 'time' in all four labels to underscore its polysemy. After the first item and before the other three I briefly touch upon *the whole of time*, which is not an element or an aspect of our experience but has come to be permanently associated with it.

(i) *Waiting times*: If you insist on spoiling your *espresso* by drinking it with sugar, you must wait for the sugar to dissolve in the beverage. Much of our lives consists in waiting for one or the other thing to happen and, ultimately, of course, we always are waiting for death. We spontaneously quantify waiting times, but our estimates are rough and highly dependent on context. However, our ancestors discovered that many readily typified natural processes have equal or proportional waiting times and began using them to measure time lengths intercontextually (or "objectively", as philosophers like to say). Thus, one may presume that cavemen soon realized that they had to wait the same time for two equally sized pots of water to boil, after placing

them over like fires. With the invention and improvement of clocks, the measurement of waiting times became ever handier and eventually took pride of place in our system of life.[3]

(ii) *The whole of time*: Waiting times can be divided into smaller parts and combined into larger wholes. At sunrise I wake up waiting for the next sunset but also for the next noon. Through an effortless idealization, we combine all waiting times into a single whole, ‘‘till Kingdom come’’. We view finished times as getting somehow packed into Life’s attic. With a little imagination and a lot of abstract construction, we come to regard this storage place as reaching back to our birth, to the beginning of human history and prehistory, and even to ‘‘the creation of the world’’ (Friedmann, 1922, p. 384).

(iii) *Points in time*: The parts of time are marked and bounded by *events*. In real life even the slightest event—e.g. the quick utterance of a monosyllable—lasts for a while and thus fills a *part* of time. However, the notion of a *point in time*, which takes no time but stands between two consecutive parts of time, played a role in the arguments of Zeno of Elea and was carefully articulated by Aristotle. With the generalized use of clocks and watches this notion became an important ingredient of ordinary common sense. Indeed, to reach it one needs very little mathematical sophistication. The shadow of a vertical stick shrinks continually as the Sun climbs, attains its minimum length at noon and thereupon slowly grows again. Given these circumstances, it seems reasonable to conceive noon as a point in time, the durationless instant at which the shadow stops shrinking and begins to grow.

(iv) *Time order*: Every *waiting* time begins, goes on and usually finishes. Thus, there is an inbuilt order of succession among its parts. This is readily extended to the whole of time, for whose beginning and end most of us therefore naturally feel inclined to ask. Let us designate parts of time by lower case italics, a, b, c,…, and points in time by upper case italics A, B, C,… Then, for any two parts a and b, either (1) a is a part of b, or (2) b is a part of a, or (3) a and b share a part of time c, or (4) a and b do not have any part in common. In case (4), either a has already ended when b begins, in which case we say that a precedes b, or a begins after b has ended, in which case we say that a follows b. If a precedes b and b precedes c, then a precedes c. All this, I dare say, is fairly obvious. We have thus a linear order among non-overlapping parts of time. Points in time readily inherit this order if we make the following common though far from obvious assumption: if A and B are any two distinct points in time, there are always two non-overlapping parts of time a and b, such that A belongs to a and b belongs to B ($A \in a \wedge B \in b$). Under this assumption, a linear order is established among time points if we stipulate that, for any two such points A and B, A precedes B if and only if there are two non-overlapping parts of time a and b such that $A \in a$ and $B \in b$ and a precedes b (in the sense defined above). As far as I can tell, in every case in which I distinguish two given points in time A and B, a third point C can be discerned, such that either A precedes C and C precedes B or B precedes C and C precedes A. This familiar experience encourages one to conceive the linear order of points in time as

[3]We now know that time measurement by clocks also depends on context, insofar as their accuracy is controlled by atomic clocks, which measure actual waiting times *along their respective worldliness.* Thus, if Max remains seated on an inertially moving spaceship while his twin sister Una takes a roundtrip from it to α-Centauri, Max will wait longer than Una for their reunion, *according to their respective standard clocks.*

dense in itself. Modern mathematical physics goes a large step further and regards it as continuous, and indeed as a linear order on a differentiable manifold (more on this below).

(v) *Past, future and current time*: Perhaps the most salient feature of our life in time is the partition of events and their times of occurrence into past, present and future. As far as I can judge, every normal 4-year-old child understands this partition and regularly applies it to matters of interest to him or her. My judgment may be biased by the fact that all the 4-year-olds I have talked to spoke either Spanish or English and had already mastered the use of tenses. Kant, who also spoke an Indo-European language, once noted: "All predicates have as copula: *is*, *was*, *will be*" (Kant, 1902, 17:579; R. 4518). The partition is central to our consciousness and our behavior and most of our decisions would hardly make any sense without it. Nevertheless, it has been declared illusory by respectable thinkers.

The partition of times noted under (v) is closely linked to the four acceptances of 'time' I commented under (i)–(iv). Thus, (i) one currently waits only for future events; to wait for the past to happen, though perhaps feasible for someone who adopts "the point of view from nowhen" (Price, 1996), sounds crazy and even ungrammatical in ordinary English. Indeed our most primitive idea of a duration or length of time is how much we must wait now until an expected future event—e.g. the departure of a plane we have already boarded—becomes past. The partition naturally extends (iii) to points in time, and is linked (iv) to their time order, so that every past instant precedes all future instants. The partition provides the basic empirical criterion for establishing a time order among events: event A precedes event B if A is present or past when B is future, or A is past when B is present or future. Applied to (ii) the whole of time, the partition leads to the abstract conception of time already found in Aristotle. Using the clear and precise language of modern mathematics, we can say that, according to this conception, the whole of time is a linear continuum in which the present instant, *now* (τὸνῦν), effects a Dedekind cut. There is an obvious difficulty—a contradiction, perhaps?—in any statement that uses the word 'now' to refer to the present point in time, inasmuch as the word will no longer denote its original referent when the statement is finally completed. Some philosophers believe that this difficulty can be evaded by avoiding all mention of *now* and using instead the so-called *tenseless present*. A coordinate system is defined on the whole of time, which assigns a unique numerical label to each instant. The point in time when a particular event E occurs can then be denoted by its label: 'E occurs at time t', say. This approach has fostered the opinion that all times are homogeneous and that their partition into past, present and future is illusory. However, the tenseless present remains a mere figment of the intellect, devoid of reference, unless the t-labels are anchored to the time of our life, which, as we know too well, is structured around that partition. The zero of time of the Christian or "common" era must be fixed at so many years, days and hours before *now*, lest it should float timelessly in nowhen (cf. Auyang, 1998, p. 226).

4. The mathematical structure $\mathbb{T}$

Classical mathematical physics took an interest in most of the said connotations of 'time', which it succeeded in representing as features of a one-dimensional differentiable manifold. The concept of a differentiable manifold is, of course, a creature of the 20th

century, but the classical differential equations of physics make no sense unless the time variable that occurs in them ranges over a domain that we can bring under this concept. Classical physico-mathematical time is topologically equivalent—indeed, diffeomorphic—with the real continuum $\mathbb{R}$. However, to emphasize that its structure does not incorporate the full richness of the complete Archimedean ordered field normally denoted by the symbol $\mathbb{R}$, I shall designate it by the non-standard symbol $\mathbb{T}$.[4] Any smooth bijective mapping $f : \mathbb{R} \to \mathbb{T}$ transmits to $\mathbb{T}$ the order relations between $\mathbb{R}$'s points and the metric relations between $\mathbb{R}$'s intervals. Every such mapping f takes values $f(0)$ and $f(1)$ at the neutral elements of the additive and the multiplicative group of $\mathbb{R}$, respectively. Nevertheless, we regard any such mapping f as "forgetting" the algebraic structure of $\mathbb{R}$, for we do not attach any sense to the operation of multiplying one time interval by another.

The comparative poverty of $\mathbb{T}$ *vis-à-vis* $\mathbb{R}$ has one implication of some interest for our subject. Any smooth bijective mapping $f : \mathbb{T} \to \mathbb{R}$ is a *global time coordinate* function. Every global time coordinate t induces on $\mathbb{T}$ a linear order $<_t$ such that, for all $a, b, c \in \mathbb{R}$, we have that $t^{-1}(a) <_t t^{-1}(b) <_t t^{-1}(c)$ iff $a<b<c$. If t and t' are two global time coordinates, then the orderings induced by them on $\mathbb{T}$ either agree or are the exact reverse of each other. The latter occurs, for instance, if $t'(x) = -t(x)$ for every $x \in \mathbb{T}$. In this case, we may denote the mapping t' by $-t$ or, for greater clarity, by $(-t)$. The coordinate transformation $(-t)\circ t^{-1}$ (which maps $\mathbb{R}$ onto itself) is usually called *time reversal*, although this name would perhaps suit better the matching point transformation $(-t)^{-1}\circ t$ (which maps $\mathbb{T}$ onto itself). Now, while $(-t)^{-1}\circ t$ is an order isomorphism from $\langle \mathbb{T}, <_t \rangle$ to $\langle \mathbb{T}, <_{-t} \rangle$, the coordinate transformation $(-t)\circ t^{-1}$ is not an automorphism of the ordered field $\mathbb{R}$.[5]

$\mathbb{T}$ affords a coherent representation of four of the five "time" features of human experience I listed above. By collecting them into a single structure, $\mathbb{T}$ justifies the use of a noun 'time' that denotes any or every realization of $\mathbb{T}$. *Time points* or *instants* (iv) are naturally identified with the points of the topological space $\mathbb{T}$; *waiting times* (i) or durations with its *connected* open sets (which constitute a basis of its topology). On this understanding, *time order* (iii) necessarily agrees with one or the other of the two linear orders admitted by $\mathbb{T}$. Thus, there is apparently no problem in equating *the whole of time* (ii) with a realization of $\mathbb{T}$. On the other hand, there is nothing whatsoever in the structure of $\mathbb{T}$ that even hints at a distinction between one particular instant and the others. Moreover, the structure of $\mathbb{T}$ comprises nothing that, given the conventional choice of a particular instant, would mark an important difference between the instants that precede it and those that follow it. Indeed, there are no grounds in $\mathbb{T}$ for identifying one of its two admissible orderings and contradistinguishing it from the other. Therefore, the item (v) in my list, the trichotomy of times into past, present and future, though central to our conscious lives and crucial to our decisions, is not reflected in any way in the classical physico-mathematical representation of time.

For all practical purposes, physicists are content to use the impoverished structure $\mathbb{T}$ to set up and solve their physico-mathematical problems. When the time comes to apply and

[4]The structure $\mathbb{T}$ is specially tailored to fit Newton's "absolute, true and mathematical time" (1687, p. 5); but it also suits the universal time relative to an inertial frame defined by Einstein (1905, Section 1), as well as the mildly unprincipled domain of definition of the time coordinate that occurs in Schrödinger's equation. However, additional qualifications and caveats may be needed to speak sensibly about time in the context of General Relativity, as Gordon Belot (2007) and Butterfield and Earman (2007) has aptly noted.

[5]$(-t)\circ t^{-1}(1) = -1$, which, in contrast with 1, is not a distinguished element of $\mathbb{R}$.

to test their solutions, they put in "by hand" the link to the present and the attendant preferred time order. Indeed, physicists do this spontaneously and infallibly. If, as Einstein believed, they are yielding to a delusion,[6] it is a pretty stable and law-abiding one. I have never heard of a physicist who took what is currently going on in the lab for what went on yesterday or who failed to distinguish the outcome of an experiment from its preparation. Most working physicists accept that this is how things are and leave it at that; and sensible philosophers should presumably do the same. Nevertheless, physicalist metaphysicians expect physics to account for every major facet of their experience, and a feature so pervasive as the trichotomy of times (at each instant)—even if it is a mere illusion—cannot be an exception. They must find a physical ground, if not for the fleeting singularity of the present, then at least for the steadfast and unmistakable difference between the direction from past to future and that from future to past. That is, they must secure a physical foundation for distinguishing between the *prospective* and the *retrospective* time order. There is nothing in $\mathbb{T}$ that can represent such a foundation, but one may expect to find a suitable stand-in for it among the real-valued functions on $\mathbb{T}$ or other notional enrichments of the original structure, through which mathematical physics conceives the evolution of phenomena.

5. Time asymmetry and the laws of physics

The world we experience teems with readily discernible processes that display time-asymmetric patterns of succession. However, the craving for unity that was still so very much alive in the 19th century did not favor the dispersion of explanatory grounds over a dappled collection of sources, but would rather focus on a single unidirectional universal law of becoming, from which one would then hope to derive the entire array of temporally oriented patterns. Since the 1860s, almost all philosopher–scientists who have pursued this question have placed their stakes on the second law of thermodynamics. As popularly understood, the second law says—or implies—that there is a physical property of the universe that Clausius (1865) called *entropy*, which takes a real value at each instant and increases monotonically with time.[7] This, if true, is sufficient physical ground for distinguishing permanently and globally a definite direction on $\mathbb{T}$. However, in the 1850s

[6] Einstein wrote on 21.05.1955 to his friend Besso's widow: "Für uns gläubige Physiker, hat die Scheidung zwischen Vergangenheit, Gegenwart und Zukunft nur die Bedeutung einer wenn auch hartnäckigen Illusion" (quoted by Dorato, 1995, p. 13).

[7] "One can express fundamental laws of the Universe that correspond to the two main laws of thermodynamics in the following simple form: 1. The energy of the Universe is constant. 2. The entropy of the Universe tends to a maximum." (Clausius, 1867, p. 44, as quoted in English by Uffink, 2003, p. 129; Greven, Keller, & Warnecke, 2003). Uffink notes that in his textbook of 1876 Clausius did not include this sweeping formulation of the second law, for which he obviously did not have a shred of evidence. Nevertheless it is untiringly repeated, often with great fanfare, in the philosophical literature, e.g. by Albert (2000, p. 32): "The third and final and most powerful and most illuminating of the formulations of the second law of thermodynamics [...] is that 'the total entropy of the world (or of any isolated subsystem of the world), in the course of any possible transformation, either keeps the same value or goes up'." Indeed Brush (1976, p. 579), who says that the statement about cosmic entropy was eliminated in the *third* edition of Clausius's (1887) treatise, mentions this fact with a tinge of regret. More recently, Price (2002, pp. 88–89) has suggested that "we could do without the notion of entropy altogether" and "hence by-pass a century of discussions about how it should be defined", or perhaps use the term 'entropy' only as a portmanteau word for "a long list of the actual kinds of physical phenomena which exhibit a temporal preference, which occur in nature with one temporal orientation but not the other".

the conception of heat as a kind of motion[8] had finally prevailed over the notion that heat is a peculiar substance. By accepting that conception, physicists placed themselves under an obligation to provide a mechanical explanation of thermal phenomena, and in particular to derive the time-asymmetric second law from the time-reversal invariant laws of mechanics. This, in a nutshell, is the problem of time's arrow. A child or an Andean peasant who understood its terms would promptly conclude that it is insoluble.[9] But European adults are a stubborn breed, and some of them, from Ludwig Boltzmann on, have spent untold hours trying to figure out a solution.

Criticism of Boltzmann was promptly voiced by Loschmidt (1876), clarified by Burbury (1894) and backed—with a different argument—by Zermelo (1896a, b). Their mathematical strictures eventually compelled Boltzmann to assign a regional scope (restricted *both* in space and time!) to the direction of time resulting from the evolution of entropy. Philosophers have been surprisingly complacent about this curious view, which they have sought to bolster with schemes of their own making.[10] On the other hand, it is only very recently that HPS research, mainly by Uffink (2001, 2003; see also Brown & Uffink, 2001; Callender, 2001; Greven et al., 2003) has made it clear to philosophers that the thermodynamic concept of entropy can only be defined for particular physical systems under special conditions. This is sufficient to dismiss the popular understanding of the second law of thermodynamics as a law of cosmic evolution, to disqualify thermodynamic entropy as the physical source of universal time order, and to remove the need for deriving Time's Arrow—*per impossibile*—from the mechanical or statistico-mechanical principles of thermal physics. I cannot give here a detailed and accurate picture of this complex affair, but the following sketch is enough for my present purpose (and will, I hope, provoke a desire to read more about it in the references I give).

6. The second law of thermodynamics

The second law of thermodynamics can be traced back to Sadi Carnot's groundbreaking thoughts about heat engines (1824). A heat engine is a device by which heat is transferred from a hot reservoir—the furnace (*foyer*)—to a cooler one—the refrigerator (*réfrigerant*)—and which through this process yields mechanical work. According to the caloric theory of

[8]This phrase "the kind of motion we call heat" was introduced by Clausius (1887) in the title of one of his great papers on the subject. In our days, Stephen G. Brush used it in the title of his monumental history of the kinetic theory of heat (1976).

[9]Cf. Henri Poincaré (1893), p. 537: "Il n'est pas besoin d'un long examen pour se défier d'un raisonnement où […] l'on trouve en effet la réversibilité dans les prémisses et l'irréversibilité dans la conclusion."

[10]Here is a small sample of texts from Reichenbach (1956, pp. 127–128, my italics): "The total entropy of the world in its present state is not too high: the universe has large reserves in ordered states, so to speak, which it spends in the creation of branch systems and thus *applies to provide us with a direction of time*. […] It follows that we cannot speak of a direction for time as a whole; *only certain sections of time have directions, and these directions are not the same*. […] Boltzmann has made it very clear that the alternation of time directions represents no absurdity. He refers our time direction to that section of the entropy curve on which we are living. If it should happen that 'later' the universe, after reaching a high-entropy state and staying in it for a long time, enters into a long downgrade of the entropy curve, then, for this section, time would have the opposite direction: human beings that might live during this section would regard as positive time the transition to higher entropy, and thus their time would flow in a direction opposite to ours. […] Life is restricted to the temperate zones of transition in the entropy curve. Thus an alternation of time directions would involve no contradiction to experiences accessible to us. *Perhaps we are, indeed, inhabitants of a second section, in which the entropy 'really' goes down, without our knowing it*."

heat, which Carnot took for granted, heat is an indestructible substance, so that, if the process is carried out adiabatically, that is, in thermal isolation from the rest of the world, the amount of heat drawn from the furnace must be equal to the amount surrendered to the refrigerator. But Carnot's reasoning does not depend on this,[11] but on the assumption that the endless production of mechanical work (*création indéfinie de puissance motrice*), without consuming heat or any other agent whatsoever, is impossible (Carnot, 1824, p. 21). From this assumption, he proved that a periodically operating heat engine which in each full cycle $\mathscr{C}$ performs an amount of work $W(\mathscr{C})$ by transferring heat $Q(\mathscr{C})$ from a furnace at temperature θ^+ to a refrigerator at temperature θ^- has an efficiency $W(\mathscr{C})/Q(\mathscr{C})$ equal to or less than a maximum $C(\theta^+,\theta^-)$, and that the value of $C(\theta^+,\theta^-)$ does not depend on the nature of the means employed but "is fixed solely by the temperatures of the bodies between which the transfer of heat ultimately occurs" (p. 38). Moreover, the maximum efficiency $C(\theta^+,\theta^-)$ can only be attained if the bodies involved in the process of producing work by heat transfer do not undergo "any change of temperature which is not due to a change in volume" (p. 23). Such changes can only be effected by outside intervention on an adiabatically closed system (e.g. by moving a piston very slowly). Carnot's stated intention was to contribute with a general theory to the improvement of French technology. Eventually, his results had an effect on the design of successful heat engines, but only after William Thomson—later Lord Kelvin—took notice of it on the other side of the Channel (1848, 1849).[12]

When the caloric theory was finally given up around 1850, the amount of work W was equated with the difference $Q^+(\mathscr{C}) - Q^-(\mathscr{C})$ between the heat extracted from the furnace and the heat surrendered to the refrigerator. In the new context the efficiency is defined as $W(\mathscr{C})/Q^+(\mathscr{C})$. Rudolf Clausius and William Thomson derived Carnot's theorem (thus understood) from two differently stated "axioms", which I quote in Thomson's wording (1851; in 1882, pp. 179, 181):

> THOMSON: It is impossible, by means of inanimate material agency, to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of the surrounding objects.
> CLAUSIUS: It is impossible for a self-acting machine, unaided by any external agency to, to convey heat from one body to another at a higher temperature.

Thomson notes that, although these axioms "are different in form, either is a consequence of the other". They became known as the second law—or Principle—of Thermodynamics (energy conservation being the first).[13] Their empirical warrant is the thermal phenomena that corroborate Carnot's theorem.

[11]Carnot must have had grave doubts about the caloric theory, for his book contains the following rhetorical question: "Can one conceive the phenomena of heat and electricity as due to anything else than the motions of bodies? As such, must they not be subject to the general laws of mechanics?" (Carnot, 1824, p. 21n; cf. p. 37n).

[12]I thank Hasok Chang for pointing out to me the initial ineffectiveness of Carnot (1824). Pietro Redondi (1980) has studied the early impact of Carnot on French technology. He has found references to it in texts concerning the design of several heat engines using air (instead of steam), none of which apparently came to fruition, and also in theoretical writings by prominent engineers beginning with Clapeyron (1834). Thomson owed his acquaintance with Carnot's work to this memoir by Clapeyron.

[13]According to Thomson (1851) "the whole theory of the motive power of heat is founded on the two following propositions", viz. , "PROP. I. (Joule)", which amounts to energy conservation, and "PROP. II. (Carnot and Clausius)—If an engine be such that, when it is worked backwards, the physical and mechanical agencies in every part of its motions are all reversed, it produces as much mechanical effect as can be produced by any

"In a series of papers, Clausius and Kelvin extended and reformulated the result. In 1854 Kelvin showed that the absolute temperature scale $T(\theta)$ can be chosen such that $C(T^{+},T^{-}) = J(1-T^{+}/T^{-})$ [where J is Joule's constant], or equivalently

$$\frac{Q^{+}(\mathscr{C})}{T^{+}} = \frac{Q^{-}(\mathscr{C})}{T^{-}}.$$

Generalizing the approach to cycles involving an arbitrary number of heat reservoirs, they obtained the formulation[14]:

$$\oint_{\mathscr{C}} \frac{đQ}{T} = 0 \text{ if } \mathscr{C}, \text{is reversible} \tag{1}$$

and

$$\oint_{\mathscr{C}} \frac{đQ}{T} \leqslant 0 \text{ if } \mathscr{C} \text{ is not reversible.} \tag{2}$$

Note that here T stands for the absolute temperature of the heat reservoirs; it is only in the case of (1) that T can be equated with the temperature of the system." (Uffink, 2003, pp. 126–127; Greven et al., 2003).

Since the integral $\oint_{\mathscr{C}} đQ/T$ equals 0 whenever the body, evolving from an initial state A_0 through any series of other states, returns to A_0, the integrand "must be the total differential of a quantity that depends solely on the present state of the body and not on the way by which it has reached that state" (Clausius, 1865, Section 4). Clausius designates this quantity by S and calls it 'entropy' (*Entropie*, "from the Greek word τροπή, transformation"—Clausius, 1865). Therefore, we have that

$$dS = \frac{dQ}{T} \tag{3}$$

> or, if we suppose that this equation is integrated for a series of reversible transformations, through which the body passes from the initial state to its present state, and if we denote by S_0 the value of S for the present state, then
>
> $$S = S_0 + \int \frac{dQ}{T}. \tag{4}$$
>
> (Clausius, 1865, Section14)

(*footnote continued*)
thermodynamic engine, with the same temperatures of source and refrigerator, from a given quantity of heat" (1882, p. 178). Prop. II is then derived from the Thomson axiom quoted above, for which he argues thus: "If this axiom be denied for all temperatures, it would have to be admitted that a self-acting machine might be set to work and produce mechanical effect by cooling the sea or earth, with no limit but the total loss of heat from the earth and sea, or, in reality, from the whole material world" (1882, p. 179n.).

[14] I have renumbered the equations in Uffink's text. The symbol đ indicates that $đQ$ might not be an exact differential. 'Reversible' here translates 'umkehrbar', as defined by Clausius (1864, p. 251): a process is reversible if it proceeds so slowly that the system always remains close to equilibrium; see Uffink 2001, p. 384. (Clausius' text is given by Uffink on p. 335).

The second law can then be restated as saying (i) that the entropy of a heat engine operating under conditions of maximal efficiency remains constant in each cycle and (ii) that if the engine works under any other conditions its entropy necessarily increases.[15] However, to speak of "the entropy of the universe" as Clausius went on to do right away (see footnote 7) is only an exercise in fanciful *Naturphilosophie*. Not only did Clausius lack any empirical warrant for his cosmic version of the law. The very science of thermodynamics was wholly focused on small thermally isolated bodies whose volume and shape can be altered adiabatically by outside intervention, and the concept of temperature and the related concept of entropy were defined only for systems in a state of equilibrium which cannot be seriously ascribed to the universe as we know it. The rigorous formulation of thermodynamics, pioneered by Gibbs (1875/1878), carried out by Carathéodory (1909) and recently perfected by Lieb and Yngvason (1999) has made the second law "independent of models [...], Carnot cycles, ideal gases and other assumptions about such things as heat, temperature, reversible processes, etc." (Lieb & Yngvason, 2003, p. 147), but still defines "the additive and extensive entropy function S" only for equilibrium states.[16]

7. The kinetic theory of heat and Boltzmann's *H*-theorem

Although Clausius and Kelvin embraced the conception of heat as a kind of motion, initially they did not agree on what kind of motion it was. While Kelvin was inclined throughout his life to view matter as being ultimately continuous,[17] Clausius (1857) sought to derive the thermal behavior of gases from the hypothesis that a gas consists of "molecules", conceived as very small, perfectly elastic spheres that move freely, without interacting among themselves.[18] Clausius (1858) rectified the latter highly unrealistic assumption by making allowance for intermolecular collisions, and calculated the mean free path of a molecule. The molecular-kinetic theory of heat took a big stride forward in a paper read in September 1859 to the British Association by 28-year old James Clerk Maxwell (1860). By boldly resorting to considerations of probability (which Clausius had timidly broached) in the discussion of velocity changes in molecular collisions, Maxwell derived "the Final Distribution of Velocity among the Molecules of Two Systems acting

[15] Uffink (2003, p. 127) and Greven et al. (2003) recalls, however, that Kelvin never mentions the inequality (2) from which (ii) follows, and indeed calls Eq. (1) "the full expression of the second thermodynamic law".

[16] More significantly, perhaps, for the present discussion: the rigorous treatment of thermodynamics *excludes* very small and very large material systems from its scope. "Physically speaking a thermodynamic system consists of certain specified amounts of different kinds of matter; it might be divisible into parts that can interact with each other in a specified way. [...]. Our systems must be macroscopic, i.e., not too small. Tiny systems (atoms, molecules, DNA) exist, to be sure, but we cannot describe their equilibria thermodynamically [...]. On the other hand, systems that are too large are also ruled out *because gravitational forces become important*. [...] The conventional notions of 'extensivity' and 'intensivity' fail for cosmic bodies." (Lieb & Yngvason, 1999, p. 13; my italics; Greven et al., 2003).

[17] In the Baltimore lectures of 1884, while presenting a model of an elastic solid built from bell cranks and springs, Kelvin asserts emphatically: "The molecular constitution of solids supposed in these remarks and mechanically illustrated in our model is not to be accepted as true in nature" (Kargon & Achinstein, 1987, p. 110).

[18] Clausius says that he was inspired by Krönig (1856), whose molecular-kinetic explanation of thermal behavior assumed however that each molecule in his model moved in a direction perpendicular to one of the walls of a cubic container. The molecular theory of gases can be traced back to Daniel Bernoulli (1738, Section 10). Versions of it were put forward by Herapath (1821) and Waterston (1846), but met a generally cold reception. See Brush (1976).

on one another by any Law of Force'' (Maxwell, 1866; in Maxwell, 1890, 2: 43). This was subsequently modified by Ludwig Boltzmann (1868) and is therefore known as the Maxwell–Boltzmann distribution law. Maxwell and Boltzmann became thus the founding fathers of classical statistical mechanics.[19]

In the next few decades, Boltzmann vigorously pursued the ''reduction'' of thermal physics to classical mechanics.[20] As a part of this program, he introduced a generalized concept of entropy, which is also applicable outside states of equilibrium, and which allegedly supplied a statistico-mechanical foundation for time's arrow. The new concept turned up in connection with Boltzmann's proof that any gas, ''whatever may be the initial distribution of kinetic energy'' among its molecules, must in the long run approach the Maxwell–Boltzmann distribution and, once it is reached, keep it forever. Despite the ambitious generality of the phrase I have quoted from Boltzmann (1872, in Brush, 2003, p. 291),[21] his argument actually depends on several restrictive assumptions. Some of these are eventually relaxed or are at least declared relaxable, but others remain in place and determine the scope both of the said proof and of the ensuing demonstration that the (generalized) entropy of the gas continually increases until the gas acquires the Maxwell–Boltzmann distribution, and is constant thereafter. The following inescapable conditions are explicitly mentioned by Boltzmann:

(i) The gas consists of a large but finite number of molecules insulated and confined by rigid walls in a large but finite space R.
(ii) The molecules interact according to an unspecified law of force, which is however the same for all, it being assumed ''that the force between two material points is a function of their distance, which acts in the direction of their line of centers, and that action and reaction are equal'' (1872, in Brush, 2003, p. 279).
(iii) Interaction occurs only when the interacting molecules are very close and is therefore called ''collision'' by Boltzmann. Most of the time, however, the molecules move freely, i.e. with constant velocities along straight lines.
(iv) The probability that any particular molecule initially moves in a particular direction is the same as the probability that it moves in any other direction. (This can be formulated more precisely thus: let $\mathbf{x}$ denote the initial position of an arbitrary molecule; then, the probability that the unit vector $\dot{\mathbf{x}}/|\mathbf{x}|$ lies inside a particular solid angle α with its vertex at $\mathbf{x}$ is proportional to the size of α).
(v) The initial distribution of kinetic energies among the molecules is uniform on R. The exact meaning of this condition is explained by Boltzmann as follows: pick a connected space $r \subset R$ of any shape and unit volume; let $f(x,t)$ denote the number of

[19]Boltzmann's contribution is eloquently described by Uffink (2007, pp. 952–992) and Butterfield and Earman (2007), where one will also find abundant references for further study. See also Uffink (2004).

[20]I surround 'reduction' with shudder quotes because, contrary to many philosophers of my generation, I feel no sympathy for the idea of deriving the fullness of experience from dreams of reason. The obstacles met (and not overcome) by Boltzmann and his successors in their attempted reduction of thermodynamics to statistical mechanics are discussed with great acuity by Sklar (1993), chapter 9, especially pp. 345–373. Sklar concludes equanimously (yet, I suppose, not without irony): ''If we wish to claim that thermodynamics is reducible to statistical mechanics, we must have a subtly contrived model of reduction in mind.''

[21]As translated by Brush. Boltzmann's conclusion reads thus in German: ''Es ist somit strenge bewiesen, daß, *wie immer die Verteilung der lebendigen Kraft zu Anfang der Zeit gewesen sein mag*, sie sich nach Verlauf einer sehr langen Zeit immer notwendig der von Maxwell gefundenen nähern muß'' (Boltzmann, 1909, 1:345; I italicize the phrase in question).

molecules in r whose kinetic energy at time t is any real number in the interval $(x,x+dx)$; then, the distribution f is said to be uniform on R at time t if, for every real number x, the number $f(x,t)$ does not depend on the shape or the location of the unit volume space $r \subset R$.[22]

Uffink (2007, p. 964) and Butterfield and Earman (2007) mentions two additional assumptions that Boltzmann does not state but which he uses in his proof:

(vi) The distribution f is represented by a differentiable function, which Boltzmann also designates by f (without further ado); the number of molecules in R must therefore be large enough for this approximation to be viable.

(vii) f is allowed to vary only as a result of binary interactions; therefore the density of the gas must be low enough for n-particle collisions ($n > 2$) to be extremely rare. On the other hand, it cannot be so low that even 2-particle collisions are too infrequent for f to change.

According to Boltzmann "it is clear" that conditions (iv) and (v) will continue to hold forever, if they hold initially. I confess that I do not find this self-evident. I therefore tend to agree with Uffink when he lists the persistence in time of conditions (iv)–(vii) as a third unstated assumption (2007, p. 964). Boltzmann argues that if conditions (iv) and (v) are not met initially, they will be satisfied "after a very long time", for then "each direction for the velocity of a molecule is equally probable" (1872, in Brush, 2003, p. 267) and "each position in the gas is equivalent" (Brush, 2003, p. 268). Apparently, he thinks that one may in every case regard such "very long time" as already elapsed before whatever instant is the initial one in that case.

From these essential assumptions, plus a few other inessential ones,[23] Boltzmann is able to derive a differential equation for the distribution function $f(x,t)$, on whose left-hand side stands the partial derivative $\partial f(x,t)/\partial t$ and whose right-hand side sports a double integral. This differential equation is known as the *Boltzmann transport equation*. For brevity's sake, I shall denote $f(x,t)$ by $f_{\mathrm{MB}}(x,t)$ if $f(x,t)$ corresponds to the Maxwell-Boltzmann distribution. It can be easily shown that $\partial f_{\mathrm{MB}}(x,t)/\partial t = 0$. Thus, after the distribution f_{MB} is reached, it will never change.

[22] Condition (v) is tantamount to what Boltzmann (1964), pp. 40–41, describes as a state of *molecular disorder*.

[23] The provisional assumptions that Boltzmann invokes in his detailed proof (1872, Section I), but which he later removes or pronounces removable, include the following:

(viii) All molecules in R are monoatomic and equal to one another. In Section IV, Boltzmann extends his results to a gas consisting of polyatomic molecules of the same kind, "i.e. they all consist of the same number of mass-points, and the forces acting between them are identical functions of their relative positions" (1872, in Brush, 2003, p. 318), the mass-points or atoms being held together by a force that depends only on their mutual distance and acts along the line that joins them. Then, towards the end of Section IV, he observes that his calculation of entropy for such a polyatomic gas "can be carried out in the same way if several kinds of molecules are present in the same container" (1872, in Brush, 2003, p. 334); in this case, the total entropy of the system is equal to the sum of the entropies computed for each subsystem formed by molecules of the same kind.

(ix) The wall of the container that encloses the gas reflects the molecules like elastic spheres. Boltzmann adds: "Any arbitrary force law would lead to the same formulae. However, it simplifies the matter if we make this special assumption about the container" (1872, in Brush, 2003, p. 267).

By deft manipulation of his transport equation, Boltzmann (1872) proved that the quantity[24]

$$H = \int_0^\infty f(x,t)\left\{\log\left[\frac{f(x,t)}{\sqrt{x}}\right] - 1\right\}\mathrm{d}x \tag{5}$$

"can never increase, when the function $f(x,t)$ that occurs in the definite integral satisfies" the Boltzmann equation (1872, in Brush, 2003, p. 281). This is Boltzmann's famous (some might say "notorious") H-theorem. By virtue of it, the function H defined as in Eq. (5) in terms of any solution f of the Boltzmann equation satisfies the inequality

$$\frac{\mathrm{d}H}{\mathrm{d}t} \leqslant 0 \tag{6}$$

with equality holding if and only if $f = f_{\mathrm{MB}}$. The latter is, of course, the equilibrium case, for which alone thermodynamic entropy is defined. Boltzmann noted that precisely in this case, H is proportional to the entropy. He therefore introduced a generalized concept of entropy, which is related with H by the same proportionality factor in the non-equilibrium cases, where thermodynamic entropy is not defined. Inequality (6) entails then that the (generalized) entropy of any system to which it is applicable will increase while the distribution f differs from f_{MB} and therefore tends to become equal to f_{MB}, and that it will reach a maximum and henceforth remain unchanged as soon as $f = f_{\mathrm{MB}}$. (By virtue of the proportionality between generalized entropy and H, the former reaches its maximum when H attains its minimum.)

A glance at conditions (i)–(vii) is sufficient to persuade one that a proof based on them cannot lead to conclusions about the universe. Indeed, condition (i) alone should dispel any such illusion. But even if we manage to forget it—as so many writers on time's arrow have been able to do—we must still face condition (v), which, as Boltzmann (1964, p. 41) emphasizes, not only "is necessary to the rigor of the proof" but must be assumed in all applications of Boltzmann's equation. No region of the universe that contains, say, a star and a sizable chunk of interstellar space around it complies with condition (v). It may well be that "after a very long time" all such regions and the universe as a whole will meet this condition of uniformity, but it would be utterly reckless to assume, for the sake of the argument, that this "very long time" has elapsed already. Nevertheless, in the subsequent, at times passionate debate about the validity and meaning of the H-theorem, the major participants generally remained silent about the restrictions that Boltzmann's premises imposed on his conclusions. It was as if a goblin hidden in their minds had made them deaf and blind to anything that might threaten the satisfaction of their yen for global truth.

8. Loschmidt's reversibility objection and Boltzmann's defense

The two main objections to Boltzmann's H-Theorem are the *reversibility* objection, soon raised by Joseph Loschmidt (1876),[25] and the *recurrence* objection, due to Ernst Zermelo

[24] Boltzmann (1872) denoted this quantity by *E*. The now standard designation *H* was introduced by Burbury (1890) and adopted by Boltzmann.

[25] According to von Plato (1994), p. 85, Loschmidt's objection had already been stated by William Thomson (1874). However, all I can find in Thomson's text (as reproduced in Brush, 2003, p. 351) is a clear statement of the solid ground on which the objection rests, viz. the time reversal invariance of the laws of "abstract dynamics" (Thomson's phrase), but I do not find an argument *contra* Boltzmann.

(1896a, b).[26] I shall only discuss the former, which can be explained as follows. Consider an isolated, finite classical mechanical system consisting of N point-particles that meet the assumptions of Boltzmann's proof. The system's dynamical state Σ_0 (at initial time $\tau = 0$) is fully characterized by $3N$ position coordinates $q_1(0), \ldots, q_{3N}(0)$ and $3N$ momentum coordinates $p_1(0), \ldots, p_{3N}(0)$. If the distribution $f(0)$ differs from f_{MB}, then, according to the H-theorem, the value of H for this system must decrease to a minimum $H_{\min}$, which it reaches when $f = f_{\mathrm{MB}}$, and must remain constant thereafter. Suppose this happens at time $\tau = t$. The laws of mechanics determine exactly the position and momentum coordinates $q_1(t), \ldots, q_{3N}(t)$, $p_1(t), \ldots, p_{3N}(t)$ that characterize the state Σ_t of the system at that time. Consider now a system whose state Σ_0' at $\tau = 0$ is characterized by the coordinates $q_i'(0) = q_i(t)$, $p_i'(0) = -p_i(t)$ $(1 \leqslant i \leqslant 3N)$. According to the laws of mechanics the evolution of this system from $\tau = 0$ to $\tau = t$ is exactly the reverse of that of the system we considered first. Therefore, its state Σ_t' at $\tau = t$ will be given by $q_i'(t) = q_i(0)$, $p_i'(t) = -p_i(0)$ $(1 \leqslant i \leqslant 3N)$. Clearly, for this system, the distribution $f'(0) = f_{\mathrm{MB}}$ and the initial value of $H = H_{\min}$, whereas the distribution $f'(t) \neq f_{\mathrm{MB}}$ and the final value of H will exceed its initial value. Thus, if the H-theorem holds for our first system, then it does not hold for the second one, although this is a *bona fide* classical system that satisfies the theorem's assumptions. Therefore, if P stands for "The H-theorem is true of any system that meets conditions (i)–(vii)", then, by the familiar tautology $(P \supset \neg P) \supset \neg P$, statement $\neg P$ is plainly false.

Boltzmann must have been cut to the quick by Loschmidt's objection for, although he explained it faithfully and clearly and, in the end, essentially granted it, he described it as "an interesting sophism" and set out, without more ado, "to locate the source of the fallacy in this argument" (1877a, in Brush, 2003, p. 365). However, Boltzmann's line of defense depends entirely on the fact, apparently overlooked by Loschmidt, that some of the premises from which the H-theorem is proved are statements of probability. As a consequence of this fact, the H-theorem cannot be regarded as a universal law of nature, but only as an overwhelmingly probable statistical generalization. Thus, Boltzmann does not actually disclose a fallacy at the heart of Loschmidt's argument, but rather a colossal misunderstanding, for which Boltzmann himself was partly to blame, since he had not sufficiently emphasized the unorthodox meaning and reach of molecular-kinetic statements in his former publications.[27] To elucidate it, one usually distinguishes between the microstates and the macrostates of a mechanical system of N particles. The *microstate* of

[26]Zermelo's recurrence objection rests on a theorem by Poincaré (1890), which can be stated as follows: *In a system of mass-points under the influence of forces that depend only on position in space, any state of motion must recur infinitely many times, at least to any arbitrary degree of approximation, if the position and momentum coordinates cannot increase to infinity* (Zermelo, 1896a and Butterfield and Earman (2007) in Brush, 2003, pp. 382–383). Since H depends on the distribution f, which in turn depends on the momentum coordinates, Zermelo argued that H must therefore return infinitely many times to its initial value, contrary to the original Boltzmann claim that H decreases steadily until it reaches a minimum which it retains. According to Mackey (1992, p. 45). "Zermelo was right in his assertion that the entropy of a system whose dynamics are governed by Hamilton's equations, or any set of ordinary differential equations for that matter, cannot change", but was wrong to base his argument on Poincaré's theorem; Mackey says that Zermelo's fallacy lies on "his implicit assumption that densities (on which the behavior of entropy depends) will behave like points".

[27]Jan von Plato (1994), p. 79, believes that "Boltzmann had by 1872 already a full hand against his future critics", for he was sufficiently explicit about the statistical nature of his premises and conclusions. For a more balanced judgment concerning Boltzmann's position before and after 1876, see Uffink (2007) and Butterfield and Earman (2007), Section 4.2. No matter when Boltzmann got his full house, what Loschmidt could show against it looks to me like a straight flush.

the system at time t is identified by the exact values of the $6N$ coordinates $p_i(t)$, $q_i(t)$ ($1 \leqslant i \leqslant 3N$). The set of all such $6N$-tuples fills the system's *phase space* $\mathscr{S} \subseteq \mathbb{R}^{6N}$. A *macrostate* of the system is an open set $M \subset \mathscr{S}$ formed by microstates that share the values of certain physical quantities one may plausibly regard as macroscopic observables. Evidently, if N is the number of molecules in a mere cubic meter of gas (at normal pressure and temperature), it is absolutely impracticable to identify the microstate of our system. Therefore, molecular-kinetic theory cannot predict the evolution of microstates according to the laws of mechanics, but must rely on statistical reasoning concerning the macrostates. To get started, this kind of reasoning requires the definition of a probability measure μ on the phase space.[28] Boltzmann assumes that every conceivable microstate $\langle q_1, \ldots, q_{3N}, p_1, \ldots, p_{3N} \rangle \in \mathscr{S}$ is equally probable; judging by his reasoning, it appears that he took this to mean that μ is uniformly distributed over $\mathscr{S}$. Hence, for any open set $M \subset \mathscr{S}$, $\mu(M)$ is proportional to the Euclidean volume of M. There are, of course, no a priori grounds for this assumption, but it can somehow be justified a posteriori by the predictive success of inferences based on it. If $M_{\mathrm{MB}} \subset \mathscr{S}$ is the set of microstates characterized by the Maxwell–Boltzmann distribution f_{MB}, it is easy to show that $\mu(M_{\mathrm{MB}})$ is very large, indeed very much larger than the measure of any other macrostate $M \subset \mathscr{S}$. Therefore, according to Boltzmann, if our system is initially in a macrostate M_I for which the distribution $f \neq f_{\mathrm{MB}}$, then almost every microstate in M_I must eventually evolve into a microstate belonging to M_{MB}. This evolution will take more or less time depending on the microstate, but while the evolution lasts the function H will steadily decrease until it reaches its minimum $H_{\min}$, as f becomes equal to f_{MB}. Prompted by Loschmidt's challenge, Boltzmann (1877b, 1878) wrote a classical paper "On the relation between the second principle of the mechanical theory of heat and the probability calculus with respect to the theorems concerning thermal equilibrium", followed by "Further remarks on some problems of the mechanical theory of heat", where he gave the definition of the entropy S of a system in terms of the probability W of its mechanical state which is carved on Boltzmann's tombstone: $S = k \log W$.

From the overwhelming value of $\mu(M_{\mathrm{MB}})$ Boltzmann infers that it is enormously likely that a point in a low probability macrostate is the starting point of an evolution leading to a microstate in M_{MB}. His inference rests on the notion that the length of time that a mechanical system spends in a macrostate M is proportional to its probability $\mu(M)$. Fortunately, the present discussion does not require that we go into the foundations and difficulties of this notion.[29] We can simply accept it and yet conclude that Boltzmann's defense is powerless against Loschmidt's objection. For ease of reference, I introduce a few symbols. I shall write (i) M_0 for the particular non-equilibrium macrostate I choose for

[28] Since measure theory and the measure-theoretic approach to probability were still unborn in the 1870s, my manner of speaking here is surely anachronistic. Nevertheless, I expect it to be helpful.

[29] In his earlier writings on the subject, Boltzmann apparently based the idea that the length of time spent by the system in a macrostate is proportional to the probability of this macrostate on the so-called *ergodic* hypothesis, according to which the trajectory of the system in phase space passes through every point of the hypersurface corresponding to the system's energy. Since this is mathematically impossible—as Rosenthal (1913) and Plancherel (1913) independently proved—it has been suggested (already by Paul and Tatiana Ehrenfest in 1912 (Ehrenfest, 1959; Ehrenfest & Ehrenfest, 1912)) that Boltzmann was actually thinking of the *quasi*-ergodic hypothesis, by which the system comes as close as you wish to every point of the energy hypersurface. This is not impossible, but it does not yield the desired consequences regarding the probability distribution (Uffink, 2007, p. 960; Butterfield & Earman, 2007).

consideration, (ii) $M_{0,t}^{\rightarrow\mathrm{MB}}$ for the proper subset of M_0 formed by the starting points of phase space trajectories that leave M_0 at time $\tau = 0$ and reach M_{MB} at time $\tau = t$; and (iii) $M_{\mathrm{MB}}^{\leftarrow 0,t}$ for the set of points at which the phase trajectories initiated in $M_{0,t}^{\rightarrow\mathrm{MB}}$ reach M_{MB}. I shall denote by $\mathscr{R}$ both (iv) the transformation of the phase space $\mathscr{S} \subseteq \mathbb{R}^{6\mathrm{N}}$ defined by $\langle q_1, \ldots, q_{3N}, p_1, \ldots, p_{3N} \rangle \to \langle q_1, \ldots, q_{3N}, -p_1, \ldots, -p_{3N} \rangle$ and (v) the mapping induced by this transformation in the power set $\wp\mathscr{S}$. The use of the same symbol for designating two such mappings is of course standard; however, to avoid needless confusion, I write, as usual, $\mathscr{R}(x)$ for the value of mapping (iv) at a microstate $x \in \mathscr{S}$, but I write $\mathscr{R}M$ for the value of (v) at a set $M \subset \mathscr{S}$. Consider in particular the set $\mathscr{R}M_{\mathrm{MB}}^{\leftarrow 0,t}$. This set is obtained by reversing—*à la* Loschmidt—the velocity of each particle in each microstate comprised in $M_{\mathrm{MB}}^{\leftarrow 0,t}$. According to the laws of mechanics, the trajectories initiated in $\mathscr{R}\mathrm{M}_{\mathrm{MB}}^{\leftarrow 0,t}$ inexorably lead in a time interval of length t to states belonging to the non equilibrium macrostate $\mathscr{R}M_0$. During that time the function H increases steadily above $H_{\min}$. By a well-known theorem named after Liouville, $\mu(M_{MB}^{\leftarrow 0,t}) = \mu(M_{0,t}^{\rightarrow MB}) \leqslant \mu(M_0) \ll \mu(M_{\mathrm{MB}})$. The mapping $\mathscr{R}: \wp\mathscr{S} \to \wp\mathscr{S}$ preserves the measure μ. Therefore $\mu(\mathscr{R}M_{\mathrm{MB}}^{\leftarrow 0,t}) = \mu(M_{\mathrm{MB}}^{\leftarrow 0,t}) = (M_{0,t}^{\rightarrow\mathrm{MB}})$. Thus, among all the possible states of our mechanical system, the set of those that will evolve in time t from an equilibrium state belonging to $\mathscr{R}M_{\mathrm{MB}}^{\leftarrow 0,t}$, in which $H = H_{\min}$, toward a non-equilibrium state belonging to $\mathscr{R}M_0$, in which $H > H_{\min}$, is not a whit less probable that the set of those that will evolve in time t from the particular non-equilibrium macrostate M_0 to the equilibrium state M_{MB}, while H shrinks to $H_{\min}$. If Boltzmann's statistical reasoning is valid, it proves (*a*) that $H = H_{\min}$ for overwhelmingly long periods of time and (*b*) that, when $H > H_{\min}$, H tends with overwhelmingly great probability to return to its minimum value. Nevertheless, the time reversal invariance of the laws of mechanics makes the following conclusion inevitable: The probability of situations that will lead in any given time t from an (admittedly improbable) non-equilibrium macrostate M_0 to equilibrium, as H decreases, is precisely equal to the probability of situations that will lead in time t from a (likewise improbable) subset of the equilibrium macrostate M_{MB} (viz. $\mathscr{R}M_{\mathrm{MB}}^{\leftarrow 0,t}$) to the non-equilibrium macrostate $\mathscr{R}M_0$, while H increases. Boltzmann's appeal to statistics and probability does not rescue the H-theorem from Loschmidt's attack.

Boltzmann in effect granted this when he brought up, towards the end of his reply to Loschmidt

> a peculiar consequence of Loschmidt's theorem, namely that when we follow the state of the world into the infinitely distant past, we are actually just as correct in taking it to be very probable that we would reach a state in which all temperature differences have disappeared, as we would be in following the state of the world into the distant future. (Boltzmann, 1877a; in Brush, 2003, p. 367)

This amazing result was reasserted by Boltzmann in his later writings and became entrenched in the philosophical literature of the 20th century (see Reichenbach, 1956; Grünbaum, 1973). I can only regard it as a piece of intellectual bravado, for which Boltzmann could not claim the faintest empirical support.[30] He probably thought that

[30]A few lines further on Boltzmann adds: "Perhaps this reduction of the second law to the realm of probability makes its application to the entire universe appear dubious". This apparent concession to ordinary intelligence is countered at once by the following remark: "Yet the laws of probability theory are confirmed by all experiments

none would ever be forthcoming, for he dared to offer the following explanation of the ostensible time-directedness of thermal phenomena:

> One can think of the world as a mechanical system of an enormously large number of constituents, and of an immensely long period of time, so that the dimensions of that part containing our own ''fixed stars'' are minute compared to the extension of the universe; and times that we call eons are likewise minute compared to such a period. Then in the universe, which is in thermal equilibrium throughout and therefore dead, there will occur here and there relatively small regions of the same size as our galaxy (we call them single worlds) which, during the relative short time of eons, fluctuate noticeably from thermal equilibrium, and indeed the state probability in such cases will be equally likely to increase or decrease. *For the universe, the two directions of time are indistinguishable, just as in space there is no up or down. However, just as at a particular place on the earth's surface we call ''down'' the direction toward the center of the earth, so will a living being in a particular time interval of such a single world distinguish the direction of time toward the less probable state from the opposite direction (the former toward the past, the latter toward the future). By virtue of this terminology, such small isolated regions of the universe will always find themselves ''initially'' in an improbable state.* This method seems to me to be the only way in which one can understand the second law—the heat death of each single world—without a unidirectional change of the entire universe from a definite initial state to a final state. (Boltzmann, 1964, pp. 446–447; my italics)

Since the prospective and the retrospective time order are not just unequivocally labeled by an apposite conventional terminology but are also experienced (in German one would say *erlebt*) as unmistakably different, Boltzmann is telling us here that it is downright impossible for someone not just to *describe* but also to *observe* an actual decrease of entropy (or a corresponding increase of H). Should it ever happen that we are actually involved in such a process, so that, say, the entropy of our surroundings was smaller yesterday than the day before yesterday, we would *perceive* yesterday as being tomorrow, and the day before yesterday as being the day after tomorrow. This mind-boggling idea came up about the same time as the hypothesis put forward by Lorentz and FitzGerald to explain Michelson's failure to detect the relative motion of the Earth and the ether. We have here two cases in which first-rate scientists sought to overcome a flagrant conflict between theory and experience by attributing to nature some kind of systematic elusiveness. But surely Boltzmann went a long step farther than his colleagues. The Lorentz–FitzGerald contraction hypothesis only extended the scope of known forces, ascribing to them a new, hitherto unsuspected effect, which should be anyway open to ordinary experimental control. But Boltzmann postulated a radical change of language and indeed of consciousness to ensure that the phenomena of entropy decrease, which turned out to be neither impossible nor unlikely according to his statistical reasoning, remained unobserved forever.[31]

(*footnote continued*)
carried out in the laboratory'' (1877a; in Brush, 2003, p. 367). This is true, but then to extend these laws from the lab to the entire universe one would have to define a statistical ensemble of which the universe itself is an instance, and, as far as I can see, any attempt to do so cannot fail to be arbitrary.

[31] It has been pointed out to me that Boltzmann's speculation concerns the observation, from our own place in the universe, of an opposite time direction in far-away regions of the universe. Suppose, however, that we grant

Yet Boltzmann did not surrender his good sense to this fancy. In his reply to Zermelo (1896b), which contains a passage that the last quotation repeats almost verbatim, this is preceded by a warning "against placing too much confidence in the extension of our thought pictures beyond the domain of experience". Nevertheless, he adds, "with all these reservations, it is still possible for those who wish to give in to their natural impulses to make up a special picture of the universe". He proposes two such pictures: Either (α) "the entire universe finds itself at present in a very improbable state" or (β) it "is in thermal equilibrium as a whole and therefore dead" except in "relatively small regions of the size of our galaxy (which we call worlds) which, during the relatively short time of eons, deviate significantly from thermal equilibrium" (Boltzmann, 1897, in Brush, 2003, p. 416). He does admit, however, that "whether one wishes to indulge in such speculations is of course a matter of taste" (Brush, 2003, p. 417).

It is said that when Loschmidt, who was Boltzmann's colleague in Vienna, first told him that his gas would return from equilibrium to its initial non-equilibrium states if all molecular velocities are reversed, Boltzmann replied to him: "Well, you try to reverse them!" Brush (1976, p. 605), from whom I have the story, has good reasons to think it is apocryphal. But it does drive home the gist of Boltzmann's statistical approach to time asymmetric thermal phenomena. Since non-equilibrium states are inordinately improbable, it is extremely difficult to pick out in the shoreless ocean of equilibrium states the pitifully small subsets from which non-equilibrium states would be reached within a sensible length of time. Therefore, although according to the laws of mechanics it is perfectly possible for Boltzmann's function H to increase above $H_{\min}$ in a thermally isolated gas in equilibrium, for all practical purposes it is impossible to prepare an experiment in which this will happen. In several passages, Boltzmann fondly hints at this fact, with some rhetorical flourish. Yet this very fact raises a big question for him, namely: Why, if states in which $H > H_{\min}$ are so enormously improbable, is it fairly easy to pinpoint physical systems that are actually in such states and to isolate them so that they evolve in a fairly short time to an equilibrium state in which $H = H_{\min}$?

9. The low entropy Big Bang

The currently fashionable reply to this question has been chiefly promoted by the great Oxford mathematician Roger Penrose, (1979, 1989, chapter 7; 2005, chapter 27).[32] It runs as follows. In the light of astronomical evidence, the universe can be represented to a good approximation (in the large) by an expanding Friedmann–Lemaître–Robertson–Walker

(*footnote continued*)
the assertion (highlighted in my quotation from Boltzmann, 1964) that a living being in such a far-away region will nevertheless perceive the direction of time toward the less probable state as the direction toward the past, and the opposite direction as the direction toward the future. Then, nothing whatsoever could prove to us that we are not in this case, that we do not live in fact in a cosmic region that currently evolves from a more probable to a less probable state, although we perceive the contrary due to the relation between time consciousness and entropy increase proposed by Boltzmann. (So that we feel every day older and closer to death, although we are really getting younger and steadily approaching birth and conception, etc.) Like other philosophical speculations that wholly detach objective reality from subjective awareness, Boltzmann's wreaks havoc with our knowledge of the former. I suppose this is bound to happen whenever one forgets that human consciousness, despite its proneness to error and delusion, is the seat and judge of truth.

[32] Penrose's idea is unquestioningly accepted by both Price and Callender, the two parties to the debate on "the origin of time's arrow" in Hitchcock (2004).

(FLRW) model and this entails that, for some as yet unknown reason or no reason at all, when expansion began some 13,000,000,000 years ago the state of the universe as a whole was an extremely improbable one, that is, a state of extremely low Boltzmann entropy. Since then the entropy of the universe has steadily increased, but it still has a very long way to go before reaching the maximally probable state of thermal death, from which the universe can then only move away through short-lived local fluctuations. Thus, one of the two pictures of the world which according to Boltzmann (1897) are available to people who wish to give in to their "natural" metaphysical impulses, namely, the one labeled (α) above, can be assigned a definite content allegedly supported by scientific cosmology. I was greatly confused when I first read Penrose's proposal, for I admired his work on General Relativity (Penrose, 1965, 1968; Hawking and Penrose, 1970) and trusted his judgment on GR matters, but, on the other hand, I was well aware of the stringent condition of homogeneity satisfied by FLRW models. *Only* if the distribution of energy on each hypersurface of simultaneity is absolutely uniform, do the Einstein field equations admit a Friedmann–Lemaître solution. I took this to mean that FLRW models *arise* and *remain* in a state of thermal equilibrium. Indeed, in the peculiar hybrid of GR gravitational theory with non-GR particle physics known as Big Bang cosmology, "the matter (including radiation) in the early stages appears to have been completely thermalized (at least so far as this is possible, compatibly with the expansion)", for "if it had not been so, one would not get correct answers for the helium abundance" (Penrose, 1979, p. 611). Indeed, the perfect thermal equilibrium between parts of the Big Bang universe which lie outside each other's horizon and therefore have never had an opportunity of interacting among themselves was initially one of the motivations of inflationary cosmology.[33] Of course, the universe *is not* in a state of thermal equilibrium and can be regarded as an expanding FLRW universe only through substantial simplification and idealization. Nevertheless, Penrose's claim that the world began in a state of very low entropy, based on a simplified and idealized world model that presupposes perfect uniformity, seemed to me baffling (to say the least).

Penrose argues that in Big Bang cosmology the early universe is in a very low entropy state due to the somewhat anomalous behavior of gravity with regard to entropy[34]:

> In many cases in which gravity is involved, a system may behave as though it has a negative specific heat.[35] [...] *This is essentially an effect of the universally attractive nature of the gravitational interaction*. As a gravitating system 'relaxes' more and more, velocities increase and the sources clump together—instead of uniformly spreading throughout space in a more familiar high-entropy arrangement. With other types of force, their attractive aspects tend to saturate (such as with a system bound electromagnetically), but this is not the case with gravity. Only non-gravitational forces can prevent parts of a gravitationally bound system from collapsing further inwards as the system relaxes. Kinetic energy itself can halt collapse only temporarily. In the absence of significant non-gravitational forces,

[33] This is nicely explained by Guth (1997), pp. 180–186.

[34] At first sight the appeal to gravity is perplexing, for, as I pointed out in footnote 16, thermodynamics cannot be rigorously applied to a material system exposed to powerful gravitational action. But surely Penrose is not talking here about thermodynamic entropy, but about "some analytic quantity, usually involving expressions such as $-p \ln p$, that appears in information theory, probability theory and statistical mechanical models" (Lieb & Yngvason, 1998, p. 571).

[35] Penrose cites the case of black holes, which get hotter as they emit Hawking radiation, and of satellites in orbit around the Earth, which speed up, rather than slowing down, due to frictional effects in the atmosphere.

when dissipative effects come further into play, clumping becomes more and more marked as the entropy increases. Finally, maximum entropy is achieved with collapse to a black hole. (Penrose, 1979, p. 612; my italics)

If gravity is naturally attractive and gravitational sources tend to clump together, the thoroughly uniform, clump-free universe assumed by relativistic cosmology (Einstein, 1917, 1987; Friedmann, 1922, 1924) may well be said to be in a very improbable state (not just initially, though, but throughout its entire history). Now, attraction was always the distinguishing character of gravity, e.g. in Aristotelian physics, and under Newton's dispensation attraction became universal. But Friedmann showed, to Einstein's initial dismay, that the gravitational interaction governed by the GR field equations can be either attractive or repulsive (even if the cosmological constant $\Lambda = 0$), depending on the parameters of the model. Thanks to Friedmann's mathematical discovery one could contemplate a viable scientific cosmology, which then came to fruition thanks to the physical discoveries of Hubble and of Penzias and Wilson. But neither the mathematics of GR nor all the splendid data of modern telescopy and radio-telescopy assign a probability to the initial state of an expanding FLRW universe (as compared, say, with the single-clump universe predicted by some forms of Newtonian cosmology). We encounter here the difficulty I already mentioned in footnote 30. The world teems with events, processes and situations to which the concept of probability and statistical reasoning are fruitfully applied; but to estimate the probability of the universe as whole involves, I dare say, a category mistake. Even if I am wrong, to set up the required terms of comparison—to define the probability space in which the initial state of our present universe is a point (or would it take up a region?)—clearly demands a much greater exertion of the human fancy than it is reasonable to allow in science. For a thorough demystification of the low entropy Big Bang I refer the reader to Earman (2006), a perspicuous and compelling example of powerful HPS criticism. Its timely publication allows me to put here an end to our trek.[36]

10. Concluding remarks

Let us recapitulate what we have done. *First* I reviewed the meanings and uses of 'time' and proposed a moderately systematic summary of those I judged relevant in the present context. I noted that theoretical physics excludes from its mathematical representation of time the trichotomy of times *now*, which we can all take stock of whenever we are awake. Suppressing it serves the scientist's pursuit of universality, but does not favor the treatment of our problem, for it leaves out the one feature of experience by virtue of which we can at every moment unmistakably distinguish on intrinsic grounds the prospective from the retrospective time order. *Next* I sketched the main attempts to find in the laws of physics a

[36]A fuller discussion of time-asymmetry in statistical mechanics would have to deal with work done in the wake of Krylov's scathing criticism of the traditional foundations of statistical mechanics (Krylov, 1979). A good review with abundant references will be found in Uffink (2007) and Butterfield and Earman (2007), Section 6. From my amateurish philosophical standpoint, I feel attracted by Mackey (1992), who studies the conditions for entropy increase to a maximum in closed thermodynamic systems and concludes that this is possible only under the special condition called *mixing*, and that it is necessary only under the more special condition of *exactness*, which can never hold in an *invertible* dynamical system such as those governed by Hamilton's equations. But I do not presume to pass judgment over these mathematical results. However, I was glad to note that the very notions of *invertibility* and *non-invertibility*—as defined by Mackey (1992, pp. 23f.) for Markov operators—*presuppose* the choice of a preferred direction of time.

sufficient reason for distinguishing one direction of time from the other. In the light of the foregoing philosophical analysis, this history of failure is not surprising, for the efforts described aimed at recovering what had been left out without actually putting it back in place. Our historico-critical exercise can thus be regarded as a pincer movement by which the "P" and the "H" wings of HPS pin down the problem of time's arrow and contrive to dissolve it. Insofar as this is a scientific problem, the foregoing discussion would corroborate Hasok Chang's claim that HPS can contribute to science. Chang says it does so "by other means", whereby he suggests that historical and philosophical methods are in and of themselves foreign to science. This suggestion will seem objectionable to anyone who recalls that some of the most decisive contributions of Newton and Einstein to physics rested squarely on their critical analysis of received notions. On the other hand, under the current regime of epistemic specialization and academic departmentalization the pigeonholing of researchers is a fact of life. Today most of those who have a standing as scientists are neither trained in HPS methods of textual and conceptual criticism nor take an interest in them. However, these methods, wielded by professional practitioners, can be useful and even necessary for certain scientific purposes. Therefore Chang's notion that HPS is "a continuation of science by other means", if generally accepted, would be of great practical value as a guide for officials who must decide on the funding of research projects.

References

Albert, D. Z. (2000). *Time and chance*. Cambridge, MA: Harvard University Press.

Auyang, S. (1998). *Foundations of complex-system theories in economics, evolutionary biology, and statistical physics*. Cambridge: Cambridge University Press.

Belot, G. (2007). The representation of time and change in mechanics In Butterfield, J., & Earman, J., (Eds.), *Handbook of the philosophy of science: Philosophy of physics* (pp. 133–227). Amsterdam: North-Holland.

Boltzmann, L. (1868). Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkte. *Sitzungsberichte der Akademie der Wissenschaften zu Wien, mathematisch-naturwissenschaftliche Klasse*, *58*, 517–560. Reprinted in Boltzmann (1909), Vol. 1, pp. 9–33.

Boltzmann, L. (1872). Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Sitzungsberichte der Akademie der Wissenschaften zu Wien, mathematisch-naturwissenschaftliche Klasse,*, *66*, 275–370 Reprinted in Boltzmann (1909), Vol. 1, pp. 316–402. English translation in Brush (2003), pp. 262–349.

Boltzmann, L. (1877a). Bemerkungen über einige Probleme der mechanischen Wärmetheorie. *Sitzungsberichte der Akademie der Wissenschaften zu Wien, mathematisch-naturwissenschaftliche Klasse*, *75*, 62–100 Reprinted in Boltzmann (1909), Vol. 2, pp. 116-122. Partial English translation in Brush (2003), pp. 362-376 (note that the title ascribed to this paper in Brush (2003), pp. vi, 362 and 362n., belongs in fact to Boltzmann 1877b).

Boltzmann, L. (1877b). Über die Beziehung eines allgemeinen mechanischen Satzes zum zweiten Satze der Wärmetheorie. *Sitzungsberichte der Akademie der Wissenschaften zu Wien, mathematisch-naturwissenschaftliche Klasse*, *76*, 373–435 Reprinted in Boltzmann (1909), Vol. 2, pp. 164–223.

Boltzmann, L. (1878). Weitere Bemerkungen über einige Probleme der mechanischen Wärmetheorie. *Sitzungsberichte der Akademie der Wissenschaften zu Wien, mathematisch-naturwissenschaftliche Klasse*, *78*, 7–46 Reprinted in Boltzmann (1909), 2: 250–288.

Boltzmann, L. (1897). Zu Hrn. Zermelo's Abhandlung »Über die mechanische Erklärung irreversibler Vorgänge«. *Annalen der Physik*, *60*, 392–398 Reprinted in Boltzmann (1909), Vol. 3, pp. 579–584. English translation in Brush (2003), pp. 412–419.

Boltzmann, L. (1909). *Wissenschaftliche Abhandlungen, herausgegeben von Fritz Hasenöhrl* (3 Vols.). Leipzig: Barth.

Boltzmann, L. (1964). *Lectures on gas theory* (Translated by Stephen G. Brush). Berkeley CA: University of California Press. German original: Vorlesungen über Gastheorie, Leipzig: Barth, 1896–1898.

Brown, H., & Uffink, J. (2001). The origins of time-asymmetry in thermodynamics: The minus first law. *Studies in History and Philosophy of Modern Physics*, *32*, 525–538.

Brush, S. G. (1976). *The kind of motion we call heat: A history of the kinetic theory of gases in the 19th century*, Vol. 2. Amsterdam: North-Holland.
Brush, S. G. (2003). In N. S. Hall (Ed.), *The kinetic theory of gases: An anthology of classic papers with historical commentary*. London: Imperial College Press.
Burbury, S. H. (1890). On some problems in the kinetic theory of gases. *Philosophical Magazine*, *30*, 301–317.
Burbury, S. H. (1894). Boltzmann's minimum theorem. *Nature*, *51*, 78–79.
Butterfield, J., & Earman, J. (Eds.). (2007). *Handbook of the philosophy of science: Philosophy of physics*. Amsterdam: North-Holland.
Callender, C. (2001). Thermodynamic asymmetry in time. *Stanford encyclopedia of philosophy* ⟨http://www.plato.stanford.edu/entries/time-thermo⟩.
Carathéodory, C. (1909). Untersuchungen über die Grundlagen der Thermodynamik. *Mathematische Annalen*, *67*, 355–386.
Carnot, S. (1824). *Réflexions sur la puissance motrice du feu et sur les machines propres a développer cette puissance*. Paris: Chez Bachelier (Facsimile reprint: Paris, Jacques Gabay, 1990.).
Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. Oxford: Oxford University Press.
Clapeyron, B.-P.-É. (1834). Mémoire sur la puissance motrice de la chaleur. *Journal de l'École Polytechnique*, *14*, 153–190.
Clausius, R. (1857). Über die Art der Bewegung, welche wir Wärme nennen. *Annalen der Physik*, *100*, 353–380 English translation in Brush (2003), pp. 111–134.
Clausius, R. (1858). Über die mittlere Länge der Wege, welche bei Molecularbewegung gasförmigen Körper von den einzelnen Molecülen zurückgelegt werden, nebst einigen anderen Bemerkungen über die mechanischen Wärmetheorie. *Annalen der Physik*, *105*, 239–258 English translation in Brush (2003), pp. 135–147.
Clausius, R. (1864). *Abhandlungen über die mechanische Wärmetheorie, Bd. 1*. Braunschweig: Vieweg.
Clausius, R. (1865). Über verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie. *Viertelsjahrschrift der naturforschenden Gesellschaft (Zürich)*, *10*, 1–59.
Clausius, R. (1867). *Abhandlungen über die mechanische Wärmetheorie, Bd. 2*. Braunschweig: Vieweg.
Clausius, R. (1887). *Die mechanische Wärmetheorie, 3. umgearbeitete und vervollständigte Auflage*. Braunschweig: Vieweg.
Dorato, M. (1995). *Time and reality: Spacetime physics and the objectivity of temporal becoming*. Bologna: CLUEB.
Earman, J. (2006). The 'past hypothesis: Not even false. *Studies in History and Philosophy of Modern Physics*, *37*, 399–430.
Eddington, A. S. (1929). *The nature of the physical world*. The Gifford lectures 1927. Cambridge: Cambridge University Press.
Ehrenfest, P. (1959). In M. J. Klein (Ed.), *Collected scientific papers*. Amsterdam: North-Holland.
Ehrenfest, P., & Ehrenfest, T. A. (1912). Begriffliche Grundlagen der statistischen Auffassung in der Mechanik. In *Encyklopädie der mathematischen Wissenschaften*, IV 2, II, Heft 6, Leipzig: Teubner. Reprinted in Ehrenfest (1959), pp. 213–302.
(1905) [Reprinted with editorial notes in Einstein (1987), Vol. 2, pp. 275–310]. *Annalen der Physik*, *17*(4), 891–921.
Einstein, A. (1917). Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie. *K. Preussische Akademie der Wissenschaften, Sitzungsberichte* (pp. 142–152). Reprinted with editorial notes in Einstein (1987), Vol. 6, pp. 540–552.
Einstein, A. (1987). *The collected papers of Albert Einstein*. Princeton, NJ: Princeton University Press.
Friedmann, A. (1922). Über die Krümmung des Raumes. *Zeitschrift für Physik*, *10*, 377–386.
Friedmann, A. (1924). Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift für Physik*, *21*, 326–332.
Gell-Mann, M., & Hartle, J. B. (1994). Time symmetry and asymmetry in quantum mechanics and quantum cosmology. In Halliwell, J. J., Pérez-Mercader, J., & Zurek, W. H. (Eds.), *Physical origins of time asymmetry* (pp. 311–345.) Cambridge: Cambridge University Press.
Gibbs, J. W. (1875/1878). On the equilibrium of heterogeneous substances. *Transactions of the Connecticut Academy of Arts and Sciences*, Vol. 3 (pp. 108–248, 343–524). Reprinted in The scientific papers of J. Willard Gibbs, Woodbridge CT: Ox Bow Press, 1993, Vol. I, pp. 55–353.
Greven, A., Keller, G., & Warnecke, G. (Eds.). (2003). *Entropy*. Princeton, NJ: Princeton University Press.
Grünbaum, A. (1973). *Philosophical problems of space and time* (second enlarged ed.). Dordrecht: Reidel.
Guth, A. H. (1997). *The inflationary universe: The quest for a new theory of cosmic origins*. Reading, MA: Addison-Wesley.

Halliwell, J. J., Pérez-Mercader, J., & Zurek, W. H. (Eds.). (1994). *Physical origins of time asymmetry*. Cambridge: Cambridge University Press.

Hawking, S. W., & Penrose, R. (1970). The singularities of gravitational collapse and cosmology. *Royal Society of London Proceedings A*, *314*, 529–548.

Herapath, J. (1821). A mathematical inquiry into the causes, laws and principal phaenomena of heat, gases, gravitation, etc. *Annals of Philosophy*, *1* (2), 273–293, 340–351, 401–406.

Hitchcock, C. (Ed.). (2004). *Contemporary debates in philosophy of science*. Oxford: Blackwell.

Kant, I. (1902). *Gesammelte Schriften, herausgegeben von der Preußischen, bzw. Deutschen Akademie der Wissenschaften, und der Akademie der Wissenschaften zu Göttingen*. Berlin: Reimer.

Kargon, R., & Achinstein, P. (Eds.). (1987). *Kelvin's Baltimore lectures and modern theoretical physics: Historical and philosophical perspectives*. Cambridge, MA: The MIT Press.

Krönig, A. K. (1856). Grundzüge einer Theorie der Gase. *Annalen der Physik*, *99*(2), 315–322.

Krylov, N. S. (1979). *Works on the foundations of statistical physics* (Translated by Migdal, A. B., Sinai, Ya. G., & Zeeman, Yu. L.). Princeton, NJ: Princeton University Press.

Lieb, E. H., & Yngvason, J. (1998). A guide to entropy and the second law of thermodynamics. *Notices of the American Mathematical Society*, *45*, 571–581.

Lieb, E. H., & Yngvason, J. (1999). The physics and mathematics of the second law of thermodynamics. *Physics Reports*, *314*, 1–96; 669.

Lieb, E. H., & Yngvason, J. (2003). The entropy of classical thermodynamics. In Greven, A., Keller, G., & Warnecke, G. (Eds.), *Entropy* (pp. 147–195). Princeton, NJ: Princeton University Press.

Loschmidt, J. (1876). Über den Zustand des Wärmegleichgewichtes eines Systemes von Körpern mit Rücksicht auf die Schwerkraft. *Sitzungsberichte der Akademie der Wissenschaften zu Wien, mathematisch-naturwissenschaftliche Klasse*, *73*, 128–142.

Mackey, M. C. (1992). *Time's arrow: The origins of thermodynamic behavior*. New York: Springer.

Maxwell, J. C. (1860). Illustrations of the dynamical theory of gases. In Maxwell (1890), Vol. 1, pp. 379–409. Originally published in *Philosophical Magazine*, *19*, 19–32; 20, 21–37.

Maxwell, J. C. (1866). On the dynamical theory of gases. In Maxwell (1890), Vol. 2, pp. 26–78. (Received on May 16 and read on May 31, 1866 at the Royal Society of London; published in the Philosophical Transactions of the Royal Society, 157).

Maxwell, J. C. (1890). In W. D. Niven (Ed.), *The scientific papers of James Clerk Maxwell*, Vol. 2. Cambridge: Cambridge University Press (Reprint: Mineola NY, Dover, 2003).

Penrose, R. (1965). Gravitational collapse and space-time singularities. *Physical Review Letters*, *14*, 57–59.

Penrose, R. (1968). Structure of space–time. In C. DeWitt, & J. A. Wheeler (Eds.), *Battelle Rencontres, 1967 Lectures in mathematics and physics* (pp. 121–235). New York: Benjamin.

Penrose, R. (1979). Singularities and time-asymmetry. In S. W. Hawking, & W. Israel (Eds.), *General relativity: An Einstein centenary survey* (pp. 581–638). Cambridge: Cambridge University Press.

Penrose, R. (1989). *The emperor's new mind: Concerning computers, minds, and the laws of physics*. Oxford: Oxford University Press.

Penrose, R. (2005). *The road to reality: A complete guide to the laws of the universe*. New York: Alfred A. Knopf.

Plancherel, M. (1913). Beweis der Unmöglichkeit ergodischer mechanischer Systeme. *Annalen der Physik*, *42*, 1061–1063 English translation in Brush (2003), pp. 521–523.

Poincaré, H. (1890). Sur le problème des trois corps et les équations de la dynamique. *Acta Mathematica*, *13*, 1–270 English translation of extracts relevant to Zermelo's recurrence argument in Brush (2003), pp. 368–376.

Poincaré, H. (1893). Le méchanisme et l'expérience. *Revue de métaphysique et de morale*, *1*, 534–537.

Price, H. (1996). *Time's arrow and Archimedes' point*. New York: Oxford University Press.

Price, H. (2002). Boltzmann's time bomb. *British Journal for the Philosophy of Science*, *53*, 83–119.

Redondi, P. (1980). *L'accueil des idées de Sadi Carnot et la technologie française de 1820 à 1860—de la légende à l'histoire*. Paris: Vrin.

Reichenbach, H. (1956). In M. Reichenbach (Ed.), *The direction of time*. Berkeley, CA: University of California Press.

Rosenthal, A. (1913). Beweis der Unmöglichkeit ergodischer Gassysteme. *Annalen der Physik*, *42*, 796–806 English translation in Brush (2003), pp. 513–520.

Sklar, L. (1993). *Physics and chance: Philosophical issues in the foundations of statistical mechanics*. Cambridge: Cambridge University Press.

Thomson, W. (Lord Kelvin) (1848). On an absolute thermometric scale founded on Carnot's theory of the motive power of heat, and calculated from Regnaults's observations. In Thomson (1882), pp. 100–106. Originally published in *Cambridge Philosophical Scoiety Proceedings*, June 5, 1848.

Thomson, W. (Lord Kelvin) (1849). An account of Carnot's theory of the motive power of heat, with numerical results deduced from Regnaults's experiments on steam. In Thomson (1882), pp. 113–155. Originally published in *Royal Society of Edinburgh Transactions*, *XVI*, 1849.

Thomson, W. (Lord Kelvin) (1851). On the dynamical theory of heat, with numerical results deduced from Mr Joule's equivalent of a thermal unit, and M. Regnault's observations on steam. In Thomson (1882), pp. 174–323. Originally published in *Royal Society of Edinburgh Transactions*, March 1851.

Thomson, W. (Lord Kelvin) (1874). The kinetic theory of the dissipation of energy. In Brush (2003), pp. 350–361. Originally published in *Royal Society of Edinburgh Proceedings*, *8*, 325–334.

Thomson, W. (Lord Kelvin) (1882). *Mathematical and physical papers collected from different scientific periodicals from May, 1841, to the present time* (Vol. I). Cambridge: At the University Press.

Uffink, J. (2001). Bluff your way in the second law of thermodynamics. *Studies in History and Philosophy of Modern Physics*, *32*, 305–394.

Uffink, J. (2003). Irreversibility and the second law of thermodynamics. In Greven, A., Keller, G., & Warnecke, G. (Eds.), *Entropy* (pp. 121–146). Princeton, NJ: Princeton University Press.

Uffink, J. (2004). Boltzmann's work in statistical physics. *Stanford encyclopedia of philosophy* ⟨http://plato.stanford.edu/entries/statphys-Boltzmann/⟩.

Uffink, J. (2007). Compendium of the foundations of statistical physics. In Butterfield, J., & Earman, J. (Eds.), *Handbook of the philosophy of science: Philosophy of physics* (pp. 924–1074). Amsterdam: North-Holland.

von Plato, J. (1994). *Creating modern probability: Its mathematics, physics and philosophy in historical perspective*. Cambridge: Cambridge University Press.

Waterston, J. J. (1846). On the physics of media that are composed of free and perfectly elastic molecules in a state of motion. *Royal Society of London Philosophical Transactions*, *183A*, 5–79 (1892). Only an abstract appeared in the year of submission: *Royal Society of London Proceedings*, *5*, 604 (1846).

Zermelo, E. (1896a). Über einen Satz der Dynamik und die mechanische Wärmetheorie. *Annalen der Physik*, *57*(3), 485–494 English translation in Brush (2003), 382–391.

Zermelo, E. (1896b). Über mechanische Erklärungen irreversibler Vorgänge. *Annalen der Physik*, *59*(3), 793–801 English translation in Brush (2003), 403–411.